A condition number for the tensor rank decomposition

Nick Vannieuwenhoven

FWO / KU Leuven

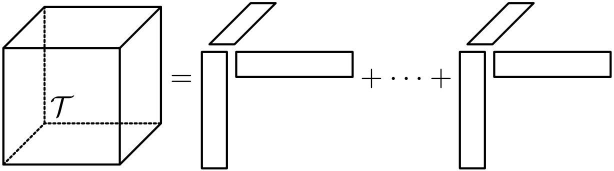
July 13, 2016

Overview

Tensor rank decomposition

Hitchcock (1927) introduced the tensor rank decomposition:[1]

$$\mathcal{T} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d$$



The rank of a tensor is the minimum number of rank-1 tensors of which it is a linear combination.

[1] Candecomp, Parafac, Canonical polyiadic, or CP decomposition.

Identifiability

A rank-1 tensor is uniquely determined up to scaling:

$$\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} = (\alpha \mathbf{a}) \otimes (\beta \mathbf{b}) \otimes (\alpha^{-1}\beta^{-1}\mathbf{c}).$$

Kruskal (1977) proved that the rank-1 terms appearing in

$$\mathcal{T} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d$$

are uniquely determined if r is small and $d \geq 3$.

Generic identifiability

It is expected[2] [BCO13, COV14] that a random real[3] or complex tensor rank decomposition

$$\mathcal{T} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d,$$

of strictly subgeneric rank, i.e.,

$$r < \frac{n_1 n_2 \cdots n_d}{n_1 + \cdots + n_d - d + 1},$$

is identifiable with probability 1, provided that it is not one of the exceptional cases where $(n_1, n_2, \ldots, n_d)$ is

- (n_1, n_2), or
- $(4,4,3)$, $(4,4,4)$, $(6,6,3)$, $(n,n,2,2)$, $(2,2,2,2,2)$, or
- $n_1 > \prod_{i=2}^d n_i - \sum_{i=2}^d (n_i - 1)$ (unbalanced).

[2] [COV14] proved the conjecture when $n_1 n_2 \cdots n_d \leq 17500$.

[3] TBA.

Perturbations and conditioning

Uniqueness is of central importance in applications, e.g., fluorescence spectroscopy, blind source separation, and parameter identification in latent variable models.

It is uncommon to work with the “true” tensor $\mathcal{T}$. Usually we only have some approximation $\widehat{\mathcal{T}}$. This discrepancy can originate from many sources:

- measurement errors,
- model errors, and
- accumulation of round-off errors.

Perturbations and conditioning

A true decomposition

$$\mathcal{T} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d$$

is nice, but I only know $\widehat{\mathcal{T}}$. I can compute an approximation

$$\widehat{\mathcal{T}} \approx \sum_{i=1}^{r} \widehat{\mathbf{a}}_i^1 \otimes \widehat{\mathbf{a}}_2^d \otimes \cdots \otimes \widehat{\mathbf{a}}_i^d \approx \mathcal{T}$$

but what does it tell me about $\mathcal{T}$?

- Is $\mathcal{T}$'s decomposition unique?
- Are the terms in $\widehat{\mathcal{T}}$'s decomposition related to those of $\mathcal{T}$?
- Can I find an upper bound on this difference?

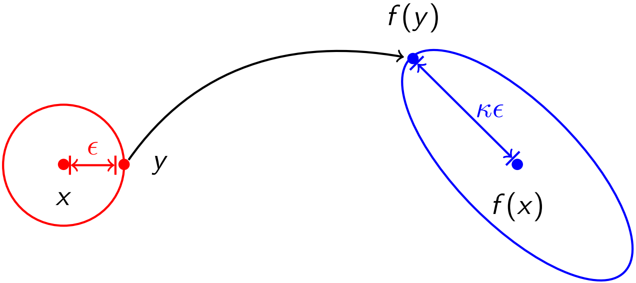
Condition number

Definition

The *relative condition number* of a function $f : X \to Y$ at $x \in X$ is

$$\kappa = \lim_{\epsilon \to 0} \max_{\|\Delta x\|_\beta \le \epsilon} \frac{\|f(x) - f(x+\Delta x)\|_\alpha / \|f(x)\|_\alpha}{\|\Delta x\|_\beta / \|x\|_\beta},$$

for some norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$.



Example

Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{bmatrix}$, $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{bmatrix}$, and

$$C_\epsilon = \begin{bmatrix} \mathbf{c} + \epsilon \mathbf{c}_1 & \mathbf{c} + \epsilon \mathbf{c}_2 & \mathbf{c} + \epsilon \mathbf{c}_3 & \mathbf{c} + \epsilon \mathbf{c}_4 \end{bmatrix}.$$

be 7×4 matrices with $\mathbf{a}_i$, $\mathbf{b}_i$, and $\mathbf{c}_i$ random vectors.

Consider a sequence of tensors

$$\mathcal{T}_\epsilon = \sum_{i=1}^{4} \mathbf{a}_i \otimes \mathbf{b}_i \otimes (\mathbf{c} + \epsilon \mathbf{c}_i) \quad \xrightarrow{\epsilon \to 0} \quad \left(\sum_{i=1}^{4} \mathbf{a}_i \otimes \mathbf{b}_i \right) \otimes \mathbf{c}.$$

Then,

- $\mathcal{T}_\epsilon$ is 4-identifiable if $\epsilon \neq 0$, while
- $\mathcal{T}_0$ has ∞-many decompositions.

Example

Let us compute the unique decomposition of $\mathcal{T}_\epsilon$ in Tensorlab using an algebraic algorithm [dL06]:

```
T_eps = cpdgen({A,B,C_eps});
[U, out] = cpd_gevd(T_eps,4);
```

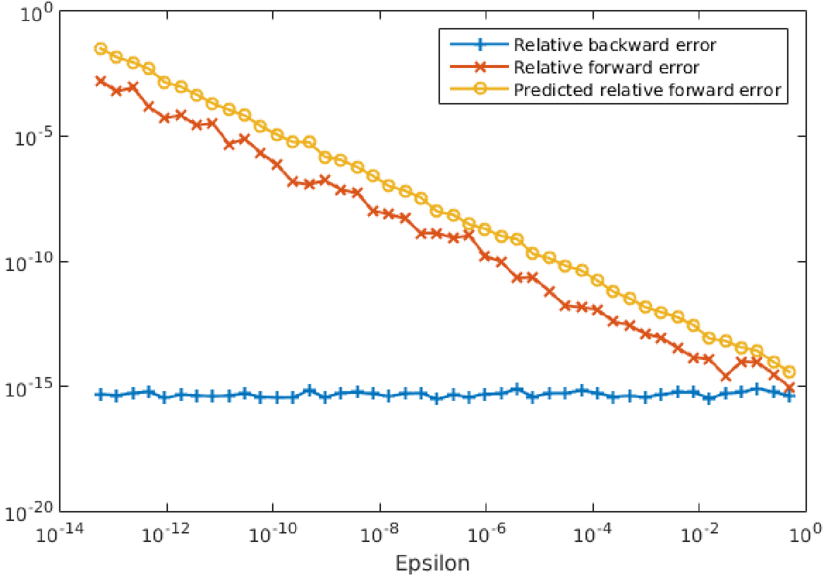
Performance measures:

- Relative backward error: $\|\widehat{\mathcal{T}}_\epsilon - \mathcal{T}\|_F / \|\mathcal{T}\|_F$.
- Squared relative forward error:

$$\frac{\|\widehat{A} - A\|_F^2 + \|\widehat{B} - B\|_F^2 + \|\widehat{C}_\epsilon - C_\epsilon\|_F^2}{\|A\|_F^2 + \|B\|_F^2 + \|C_\epsilon\|_F^2}.$$

after “fixing” the scaling and permutation indeterminacies.

Example



Relative backward error
Relative forward error
Predicted relative forward error
10^{0}
10^{-5}
10^{-10}
10^{-15}
10^{-20}
10^{-14}
10^{-12}
10^{-10}
10^{-8}
10^{-6}
10^{-4}
10^{-2}
10^{0}
Epsilon

Rough derivation: Linear approximation

Let f be the usual (overparameterized) tensor computation function:

$$f : (\mathbb{F}^{n_1} \times \cdots \times \mathbb{F}^{n_d})^{\times r} \to \mathbb{F}^{n_1 \cdots n_d}$$
$$((\mathbf{a}_1^1, \ldots, \mathbf{a}_1^d), \ldots, (\mathbf{a}_r^1, \ldots, \mathbf{a}_r^d)) \mapsto \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d.$$

By definition of differentiability, we can write

$$f(\mathbf{x} + \Delta) = f(\mathbf{x}) + J\boldsymbol{\Delta} + \mathcal{O}(\|\Delta\| \|r(\Delta)\|) \text{ with } \lim_{\Delta \to 0} \|r(\Delta)\| \to 0,$$

and where J is the Jacobian of f at $\mathbf{x}$.

Rough derivation: Terracini's Jacobian

For every rank-1 tensor

$$\mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d \in \mathbb{F}^{n_1 n_2 \cdots n_d},$$

we define the matrix

$$T_i = \begin{bmatrix} I_{n_1} \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d & \cdots & \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^{d-1} \otimes I_{n_d} \end{bmatrix}.$$

Then, the Jacobian of f at $\mathbf{x}$ is given by

$$J = \begin{bmatrix} T_1 & T_2 & \cdots & T_r \end{bmatrix};$$

I call it Terracini's matrix.[4]

[4] Terracini (1911) studied this Jacobian in his famous Lemma.

Rough derivation: Bounding the condition number

Continuing from

$$f(\mathbf{x}+\Delta) - f(\mathbf{x}) = J\boldsymbol{\Delta} + \mathcal{O}(\|\Delta\|\|r(\Delta)\|)$$
$$J^{+}(f(\mathbf{x}+\Delta) - f(\mathbf{x})) = \boldsymbol{\Delta} + \mathcal{O}(\|\Delta\|\|r(\Delta)\|),$$

we find

$$\|J^{+}\|_2 \geq \frac{\|\boldsymbol{\Delta}\|\big(1+\mathcal{O}(\|r(\Delta)\|)\big)}{\|f(\mathbf{x}+\Delta) - f(\mathbf{x})\|}$$

where J^{+} is a left inverse of J. Hence,

$$\|J^{+}\|_2 \geq \kappa = \lim_{\epsilon\to 0} \max_{\|\Delta\mathcal{T}\|\in\mathcal{G}_\epsilon} \frac{\|\Delta\|}{\|\Delta\mathcal{T}\|}$$

with $\mathcal{G}_\epsilon = \{\Delta\mathcal{T} \mid \|\Delta\mathcal{T}\| \leq \epsilon \text{ and } \exists\mathbf{y} : \Delta\mathcal{T} = f(\mathbf{y}) - f(\mathbf{x})\}$

Rough derivation: Terracini's matrix is not of full rank

The image of Terracini's matrix is contained in the tangent space to the smallest (semi-)algebraic set enclosing the tensors of (real) complex rank equal to r. At smooth points they coincide.

The rank of the $n_1 \cdots n_d \times r(n_1 + \cdots + n_d)$ Jacobian matrix J is at most $r(n_1 + n_2 + \cdots + n_d - d + 1)$.

Hence, the derivation is not that straightforward ...

Rough derivation: A bumpy road

Some issues:

1. Singular locus of r-secant semialgebraic set obstructs simple interpretation.
   - $\hookrightarrow$ Put assumption of robust r-identifiability.

2. Quotient of parameter space $\mathcal{P} = (\mathbb{F}^{n_1} \times \cdots \times \mathbb{F}^{n_d})^{\times r}$ with equivalence relation $\sim$ is not a metric space because the orbits of $\sim$ are not closed. Natural manifold-based framework of [BC13] is eliminated.
   - $\hookrightarrow$ Measure distances by a premetric (no symmetry and no triangle inequality).
   - $\hookrightarrow$ Prove continuity of inverse of f in this premetric.
   - $\hookrightarrow$ Bound forward error by operator norm of $J^\dagger$.
   - $\hookrightarrow$ Show that worst perturbation can be attained asymptotically.

The norm-balanced condition number

Theorem (—, 2016)

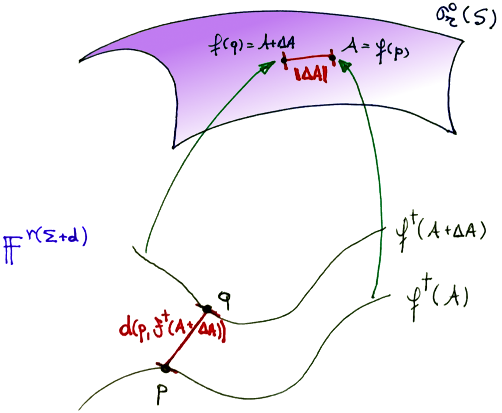
Let J be Terracini's matrix associated with the rank-1 tensors $\mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d \in \mathbb{F}^{n_1 \cdots n_d}$. Let $N = r(n_1 + \cdots + n_d - d + 1)$. If $\operatorname{rank}(J) = N$, then

$$\kappa_A = \|J^\dagger\|_2$$

is an absolute condition number of the rank decomposition problem at $\mathcal{T} = \sum_{i=1}^r \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d$ with $\|\mathbf{a}_i^1\| = \cdots = \|\mathbf{a}_i^d\|$.

Distance measure

Here's what the proposed condition number $\kappa_A(\mathbf{p})$ measures:



For every $p \in \mathbb{F}^{r(\Sigma+d)}$ and small $\|\Delta A\|$ we have:

$$d(p, f^\dagger(A+\Delta A)) \lesssim \kappa_A(p) \cdot \|\Delta A\|$$

Elementary properties

The relative condition number is scale-invariant: $\kappa(\mathcal{T}) = \kappa(\alpha\mathcal{T})$.

The condition number is orthogonally invariant.

The relative condition number of an order-d rank-1 tensor is $\sqrt{d^{-1}}$.

The case of weak 3-orthogonal tensors (—, 2016)

Let $\alpha_i \in \mathbb{R}_+$ be sorted as $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0$, and let

$$\mathcal{T} = \sum_{i=1}^{r} \alpha_i \mathbf{v}_i^1 \otimes \cdots \otimes \mathbf{v}_i^d \quad \text{with } \|\mathbf{v}_i^k\| = 1$$

be a robustly r-identifiable weak 3-orthogonal tensor:

$$\forall i < j : \exists 1 \leq k_1 < k_2 < k_3 \leq d : \langle \mathbf{v}_i^{k_1}, \mathbf{v}_j^{k_1} \rangle = \langle \mathbf{v}_i^{k_2}, \mathbf{v}_j^{k_2} \rangle = \langle \mathbf{v}_i^{k_3}, \mathbf{v}_j^{k_3} \rangle = 0,$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. Then,

$$\kappa = \alpha_r^{-1+1/d} \sqrt{\sum_{i=1}^{r} \alpha_i^2} \Big/ \sqrt{\sum_{i=1}^{r} d\alpha_i^{2/d}}.$$

If $\alpha_1 = \cdots = \alpha_r$, then $\kappa = \sqrt{d^{-1}}$.

Ill-posedness and ill-conditioning

The classic example from [dSL08] is the rank-3 tensor

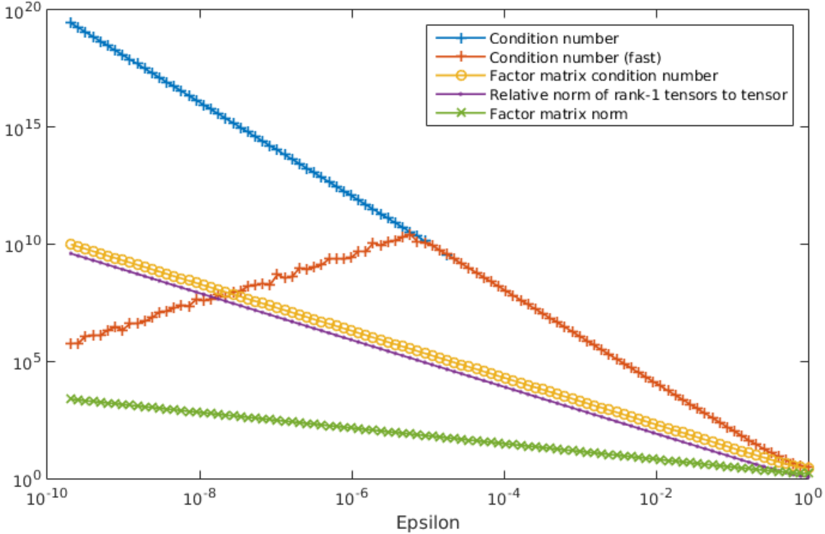
$$\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{z} + \mathbf{a} \otimes \mathbf{y} \otimes \mathbf{c} + \mathbf{x} \otimes \mathbf{b} \otimes \mathbf{c},$$

which is a limit of identifiable rank-2 tensors:

$$\lim_{\epsilon \to 0} \left(\frac{1}{\epsilon} \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} - \frac{1}{\epsilon} (\mathbf{a} + \epsilon \mathbf{x}) \otimes (\mathbf{b} + \epsilon \mathbf{y}) \otimes (\mathbf{c} + \epsilon \mathbf{z}) \right).$$

Experiments suggest that as you move towards an open part of the boundary of the r-secant variety of a Segre variety, the relative condition number becomes unbounded.

Ill-posedness and ill-conditioning



Conclusions

Take-away messages:

- Tensors are conjectured to be identifiable.
- Forward errors matter.
- The condition number multiplied with the backward error bounds the forward error to first order.
- The condition number of a decomposition can be computed practically.

Thank you for your attention!

References

Main reference:

- Vannieuwenhoven, *A condition number for the tensor rank decomposition*, arXiv:1604.00052, 2016.

Software and algorithms:

- [dL06] De Lathauwer, *A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization*, SIAM J. Matrix Anal., 2006.
- Sorber, Van Barel and De Lathauwer, *Tensorlab v3.0*, `www.tensorlab.net`.

References

Introduction

- [H1927] Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, J. Math. Phys., 1927.
- [K1977] Kruskal, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Lin. Alg. Appl., 1977.

Conditioning

- [BC13] Bürgisser and Cucker, *Condition: The Geometry of Numerical Algorithms*, Springer, 2013.
- [dSL08] de Silva and Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 2008.
- [V16] Vannieuwenhoven, *A condition number for the tensor rank decomposition*, arXiv:1604.00052.

References

Generic identifiability

- [BCO13] Bocci, Chiantini, and Ottaviani, *Refined methods for the identifiability of tensors*, Ann. Mat. Pura Appl., 2013.
- [CO12] Chiantini and Ottaviani, *On generic identifiability of 3-tensors of small rank*, SIAM J. Matrix Anal. Appl., 2013.
- [COV14] Chiantini, Ottaviani, and Vannieuwenhoven, *An algorithm for generic and low-rank specific identifiability of complex tensors*, SIAM J. Matrix Anal. Appl., 2014.
- [DdL15] Domanov and De Lathauwer, *Generic uniqueness conditions for the canonical polyadic decomposition and INDSCAL*, SIAM J. Matrix Anal. Appl., 2015.
- [HOOS15] Hauenstein, Oeding, Ottaviani, and Sommese, *Homotopy techniques for tensor decomposition and perfect identifiability*, arXiv, 2015.

References

Generic identifiability for symmetric tensors

- [COV16] Chiantini, Ottaviani, and Vannieuwenhoven, *On generic identifiability of symmetric tensors of subgeneric rank*, Trans. Amer. Math. Soc., 2016. (Accepted)
- [GM16] Galuppi and Mella, *Identifiability of homogeneous polynomials and Cremona Transformations*, arXiv, 2016.